

Functional overlap between conserved and diverged KH domains in *Saccharomyces cerevisiae* SCP160

Melissa A. Brykailo¹, Anita H. Corbett² and Judith L. Fridovich-Keil^{3,*}

¹Graduate Program in Genetics and Molecular Biology, Emory University, ²Department of Biochemistry, Emory University School of Medicine, Atlanta, GA 30322 and ³Department of Human Genetics, Emory University School of Medicine, Atlanta, GA 30322, USA

Received September 13, 2006; Revised and Accepted December 20, 2006

ABSTRACT

The K homology (KH) domain is a remarkably versatile and highly conserved RNA-binding motif. Classical KH domains include a characteristic pattern of hydrophobic residues, a Gly-X-X-Gly (GXXG) segment, and a variable loop. KH domains typically occur in clusters, with some retaining their GXXG sequence (conserved), while others do not (diverged). As a first step towards addressing whether GXXG is essential for KH-domain function, we explored the roles of conserved and diverged KH domains in Scp160p, a multiple-KH-domain-containing protein in *Saccharomyces cerevisiae*. We specifically wanted to know (1) whether diverged KH domains were essential for Scp160p function, and (2) whether diverged KH domains could functionally replace conserved KH domains. To address these questions, we deleted and/or interchanged conserved and diverged KH domains of Scp160p and expressed the mutated alleles in yeast. Our results demonstrated that the answer to each question was yes. Both conserved and diverged KH domains are essential for Scp160p function, and diverged KH domains can function in place of conserved KH domains. These findings challenge the prevailing notions about the requisite features of a KH domain and raise the possibility that there may be more functional KH domains in the proteome than previously appreciated.

INTRODUCTION

Proper gene expression depends on precisely orchestrated interactions between nucleic acids and nucleic acid-binding proteins. In fact, virtually every step of gene expression involves the activities of both sequence-specific and general nucleic acid-binding proteins. For example, transcription of DNA to RNA employs helicases,

transcription factors, RNA polymerase, 5'-capping enzymes, poly (A) polymerase and numerous other proteins required to process the nascent mRNA transcript (1,2). All of these factors must recognize nucleic acids in some manner. Then numerous mRNA-binding proteins are required to escort transcripts out of the nucleus and, ultimately, to coordinate protein translation (3). Thus, gene expression relies heavily on nucleic acid-binding proteins, with major roles for proteins that bind to mRNA. In keeping with this workload, proteins have evolved a vast array of domains that mediate interactions with RNA, including but not limited to the hnRNP K homology (KH) domain, RNA recognition motifs (RRMs), arginine-rich motifs, and zinc-finger motifs (4).

Many RNA-binding domains are abundant and highly conserved across species. In fact, Chen and Varani have estimated that about 1.5–2% of the human genome is composed of proteins that contain RRM domains, the most abundant and well-characterized of the RNA-binding domains (4). RNA-binding motifs are typically recognized and defined on the basis of their conserved primary amino acid sequence, but the degree of sequence conservation among different motifs varies widely (4). Although some major residues are defined functionally, others have been identified strictly on the basis of sequence conservation or homology (5,6). A key question is whether these conserved residues are actually critical for protein function, and therefore whether sequence conservation alone can be used as a measure of functional significance.

One remarkably versatile and highly conserved RNA-binding motif is the KH domain (7). KH domains, which were originally identified as a repeated sequence in the hnRNP K protein (7), are ~70 amino acids in length, with a characteristic pattern of hydrophobic residues, a GXXG segment and a variable loop. KH domains are often found in multiple copies per protein. There is strong evidence that KH domains are critically important to the function of those proteins that contain them. For example, fragile X syndrome, a leading cause of inherited mental retardation, can arise from a missense mutation of Ile304Asn in

*To whom correspondence should be addressed. Tel: +1 404 727 3924; Fax: 404 727 3949; Email: jfridov@emory.edu

the second KH domain of the FMRP protein (8). Indeed, the Ile304Asn mutation is associated with a severe clinical phenotype (8). Thus, there is evidence that mutations in KH domains can cause human disease, which underscores the importance of defining what constitutes a functional KH domain.

Some proteins contain both conserved KH domains that include the GXXG motif, as well as what have been termed *diverged* KH domains, in which the GXXG motif is interrupted or altered (9). Although the structures of a number of conserved KH domains have been solved (10–14), there has been little functional analysis of diverged KH domains. Recent evidence suggests possible synergistic binding to mRNA targets by the three KH domains of hnRNP K protein (15), thereby enhancing interaction with the nucleic acid substrate. Chmiel and colleagues (16) drew a similar conclusion from their work with the *Drosophila* PSI protein, which contains four KH domains. While these findings do shed light on the functional importance of KH domain *clustering* in proteins, they fail to address the functional significance of KH domain sequence conservation or divergence.

As a first step towards addressing whether GXXG is essential for KH domain function, we explored the roles of conserved and diverged KH domains in Scp160p, a multiple KH domain protein in *S. cerevisiae*. Scp160p includes 14 KH domains (17), only seven of which contain a strictly conserved GXXG motif (KH 2, 8–12 and 14). The other seven KH domains are diverged, with interruptions or alterations of the GXXG motif. Although the exact function of Scp160p and its orthologs, known as vigilins in higher eukaryotes, is not known, recent work suggests a role for this protein in modulating the metabolism of specific mRNA targets in the cytoplasm (18,19). Consistent with this postulated function, much of Scp160p is found in large protein complexes associated with soluble or membrane-bound polyribosomes (17,20–25). The goal of the present study was to explore two questions. First, are diverged KH domains essential for Scp160p function? Second, can diverged KH domains functionally replace conserved KH domains? To address these questions, we deleted and/or interchanged conserved and diverged KH domains of Scp160p, then expressed these variant alleles in yeast. We applied a combination of previously defined genetic (23) and biochemical (17,20–25) assays to the strains expressing these mutated alleles to discern the functional capacity of each encoded Scp160p protein.

Our results demonstrated that the answer to each of our two questions was yes. Both conserved and diverged KH domains *were* essential for Scp160p function, and diverged KH domains *were* capable of functioning in place of conserved KH domains.

MATERIALS AND METHODS

Yeast strains and manipulations

All yeast manipulations were performed according to standard protocols. The strains used for this study are listed in Supplemental Table 1 and, unless otherwise

noted, were derived by two-step gene replacement (26) at the *SCP160* locus from the haploid parent strain W303 [MATa *ade2-1 his3-11,15 leu2-3,112 ura3-1 trp1-1can1-100 RAD5+*, a gift from Dr. R. Rothstein, Columbia University, NY].

Construction of *FLAG.scp160* variant alleles

Unless otherwise noted, all recombinant DNA manipulations were performed according to standard procedures (27) using the XL10-Gold® (Stratagene) strain of *Escherichia coli*. To facilitate site-directed mutagenesis of *SCP160*, a 2.2 kb PstI/EcoRI fragment containing sequence from the beginning of *SCP160* KH8 through approximately 430 bp downstream of the stop codon was gel purified and ligated into pGEM3-zf(+) using T4 DNA Ligase (Invitrogen) according to the manufacturer's instructions. The resulting plasmid (JF4566, pGEM3zf+.*SCP160*.Eco.Pst) was confirmed by restriction analysis and sequencing and used as a template for the generation of all subsequent modified alleles, as detailed below.

To create *scp160* alleles individually deleted for KH13 and KH14, we used the QuikChange™ Site-Directed Mutagenesis Kit (Stratagene) according to the manufacturer's instructions, with one exception: we increased the number of PCR cycles from 18 to 28. Relevant primer sequences are listed in Supplemental Table 2. Each initial deletion was designed to remove the appropriate KH domain precisely, but also to leave in its place a short "scar sequence" that would facilitate subsequent replacement by an alternate KH sequence via gap repair homologous recombination. Primers *scpdelKH14.stu.xho.f1* and *scpdelKH14.stu.xho.r1* were used to replace KH14 with a StuI-linker-XhoI scar, resulting in the allele *FLAG.scp160Δ14.scar*. Similarly, primers *scpdelKH13.stu.xho.f2* and *scpdelKH13.stu.xho.r2* were used to replace KH13 with a StuI-linker-XhoI scar, resulting in the allele *FLAG.scp160Δ13.scar*. Finally, alleles intended to carry either a clean deletion or KH domain replacement of KH13 or KH14 were generated by gap repair of the corresponding scar construct with an appropriate fragment, as described below. The gapped backbone fragment for each recombination procedure was generated from the appropriate plasmid (pGEM3-zf(+).*FLAG.Scsp160Δ14.scar* or pGEM3-zf(+).*FLAG.Scsp160Δ13.scar*) by digestion with StuI and XhoI, followed by gel purification.

The inserts for scar replacement were generated as follows: (a) to remove the StuI-linker-XhoI scar from each deletion template, the indicated oligonucleotides in Supplemental Table 2 were annealed and extended in a single round of polymerization using the TripleMaster PCR System (Eppendorf) in a thermocycler, (b) to create *FLAG.scp160Δ14.11*, a fragment encoding KH11 was amplified from a wild-type *SCP160* template using primers *scpdelKH14.addback.KH11.f1* and *scpdelKH14.addback.KH11.r1*, (c) to create *FLAG.scp160Δ14.13*, the replacement sequence was generated by PCR using the primers *scpdelKH14.addback.13.f1* and *scpdelKH14.addback.13.r1*, (d) to create *FLAG.scp160Δ13.6*,

the replacement sequence was generated by PCR using the primers *scpd1KH13.addback.KH6.f1* and *scpd1KH13.addback.KH6.r1*, and finally (e) to create *FLAG.scp160Δ13.14*, the replacement sequence was generated by PCR using the primers *SCP.KH14.addback.f1* and *SCP.KH14.addback.r1*. All DNA fragments were gel purified using the QIAquick® Gel Extraction Kit (Qiagen), according to the manufacturer's instructions.

To accomplish gap repair homologous recombination, the appropriate purified PCR fragments were individually cotransformed into competent *E. coli* with gel-purified, gapped plasmid backbone. Transformants were cultured and the plasmids isolated using the Qiaprep Spin® Miniprep Kit (Qiagen). All resultant alleles were verified by restriction analysis and DNA sequencing across the entire region that had been generated by PCR to ensure sequence integrity.

Once the desired sequence manipulations were completed and verified, we next moved the relevant *SCP160* sequences into yeast-integrating plasmids via gap repair homologous recombination for subsequent insertion into the *SCP160* locus of the yeast genome. Towards that end, all verified plasmids were digested with *Pst*I and *Eco*RI. The ~2.2 kb band from each construct was gel purified and individually cotransformed into competent *E. coli* with pJF2147 (YIplac211.*FLAG.SCP160*) that had been gapped by digestion with *Afl*III and *Xba*I. Resultant constructs were confirmed by restriction digestion, linearized by digestion with *Pst*I, and transformed into the yeast strain JFy4493. Appropriate integration was confirmed by genomic PCR, followed by direct sequence analysis of each *SCP160* locus from KH12 to 600 bp downstream of the stop codon.

Analyses of the ability of *FLAG.SCP160* and modified alleles to complement *scp160Δ eap1Δ* synthetic lethality

We assessed the ability of all modified alleles listed in Figure 1 to function *in vivo* by quantifying the ability of each to complement *scp160Δ eap1Δ* synthetic lethality (23). Briefly, *scp160Δ eap1Δ* cells carrying both an *SCP160 URA3* "maintenance" plasmid and an *scp160* variant allele in a *LEU2* "test" plasmid were inoculated into liquid medium lacking both uracil and leucine, to ensure selection for both plasmids. Once grown to saturation, 300,000 cells were inoculated into synthetic media lacking only leucine, to enable loss of the *URA3* maintenance plasmid. When this culture reached an OD_{600} of ~2.0, again 300,000 cells were inoculated into synthetic medium lacking leucine and grown to an OD_{600} of 1.5–2.0. Finally, cells were counted, and appropriate dilutions were plated to synthetic medium lacking leucine ±5-fluoroorotic acid (5-FOA). 5-FOA is toxic only to cells that contain a functional *URA3* gene (28); therefore, the number of colonies that grew on the 5-FOA medium relative to the number of colonies that grew on medium lacking 5-FOA was a measure of the proportion of cells in the culture that had lost the *URA3* maintenance plasmid. This proportion, in turn, was a measure of the ability of the variant *scp160* test plasmid to complement loss of the

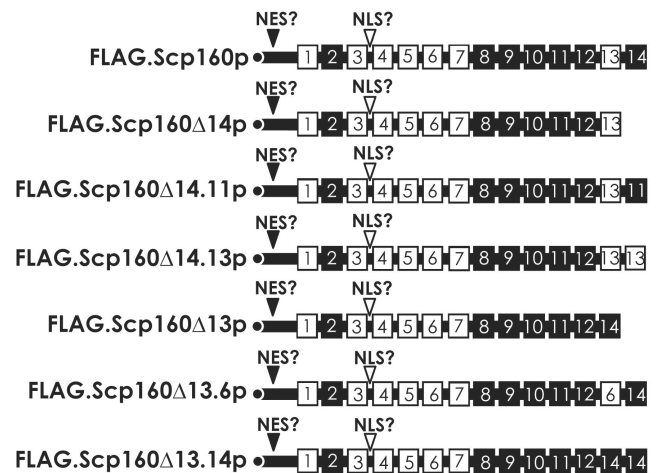


Figure 1. Predicted domain structures of wild-type and *FLAG.Scp160p* variants. Predicted domains are not drawn to scale. Numbered, filled boxes represent conserved KH domains; numbered, open boxes represent diverged KH domains. Nomenclature: the delta (Δ) symbol in each name denotes precise deletion of the indicated KH domain; a period (.) followed by a number denotes precise insertion of the indicated KH domain. Solid circles represent an N-terminal *FLAG* tag. Filled triangles represent the putative nuclear export sequence (NES); open triangles represent the putative nuclear localization sequence (NLS) (17).

wild-type *SCP160* maintenance plasmid. Empty test plasmid backbone served as the negative control, and test plasmid encoding wild-type *FLAG.Scp160p* served as the positive control. For comparison between experiments, the proportion of cells that lost the *URA3* plasmid in each strain was normalized to the corresponding proportion from the positive control strain. Therefore, the final values calculated indicated how effectively, relative to wild-type *SCP160*, each modified allele of *scp160* functioned *in vivo*. Since the *URA3* maintenance plasmid was the same in all strains and all experiments, any issues of plasmid replication efficiency or distribution, independent of *SCP160* sequence, should have cancelled out.

Biochemical analyses of *Scp160p*

All cell lyses and subcellular fractionations were performed as described previously (21,22,25). Briefly, mid-log-phase cultures were incubated with cyclohexamide at a final concentration of 100 μ g/ml for 1 min at 30°C, and then for 15 min on ice with swirling. Cells were washed in cold water, then lysed in cold buffer (25), each containing 100 μ g/ml cyclohexamide. All cell lysis procedures were accomplished using a multihead vortex mixer that was set to the highest speed at 4°C for 20 min. The supernatant from the lysis was centrifuged at 3000 rpm at 4°C for 5 min in a tabletop microfuge. The supernatant from this step was transferred to a fresh tube and centrifuged at 12000 rpm, 4°C, for 8 min in a tabletop microfuge. The supernatant from this final spin was the soluble lysate. The protein concentration of this soluble lysate was quantified using the Bio-Rad Protein Assay (cat #500-0006) according to the manufacturer's instructions. Next, 15 μ g of soluble protein from untagged

wild-type, FLAG-tagged wild-type, and FLAG-tagged Scp160p variants were each run on a gel for subsequent Western blot analysis. FLAG.Scpl60p band intensity, for the wild-type protein and all variants, was quantified by scanning densitometry, and FLAG.Scpl60p signal was normalized to the intensity of a previously reported endogenous protein also recognized by the anti-FLAG M2 antibody (21).

For analyses of polysome association, 200 μ l of soluble lysate were loaded onto a 15–45% sucrose gradient and centrifuged at 39,000 rpm in an ultracentrifuge for 2.5 h at 4°C, as described previously (25). Fractions (1 ml) were collected from the sucrose gradients and 400 μ l of each were concentrated for Western blot analysis. In brief, samples were mixed with 400 μ l MeOH and 100 μ l chloroform and agitated vigorously. Precipitated proteins were pelleted by centrifugation at 13,000 rpm for 2 min in a tabletop microfuge; the top phase was discarded, and 400 μ l MeOH was added to the pellet. Each sample was inverted to mix and then re-centrifuged at 13,000 rpm for another 5 min. This supernatant was discarded, and samples were dried in a speed vac with no heat for ~20 min until no remaining liquid was visible. The protein pellet was resuspended in 1 \times loading buffer and applied onto a 10% SDS-PAGE gel for subsequent Western blot analysis using the monoclonal ANTI-FLAG[®] M2 antibody (Sigma-Aldrich) to detect FLAG-tagged Scp160p, as described previously (25). Controls involved preincubation of lysates with either 50 units/ml RNase ONE[™] (Promega, Inc.) or 60 mM EDTA for 30 min at 4°C with gentle rocking.

For analysis of mRNP complex formation, soluble lysate from 11 of cell culture was size-fractionated over an S300 Sephacryl column, as described previously (25). Fractions of 2 ml were collected, and 12 μ l of each were analyzed via Western blot using the monoclonal ANTI-FLAG[®] M2 antibody (Sigma-Aldrich) to detect FLAG-tagged Scp160p.

RESULTS

Expression of FLAG-tagged wild-type and variant Scp160p proteins in yeast

As a first step to delineate the distinct versus potentially overlapping functions of conserved and diverged KH domains in Scp160p, we generated alleles encoding two modified variants: FLAG.Scpl60 Δ 14p, in which conserved KH14 was deleted, and FLAG.Scpl60 Δ 13p, in which diverged KH13 was deleted (Figure 1). While previous studies from our lab and others (24,25) had tested the function of Scp160p missing KH14 or KH13 + 14, the FLAG.scpl60 Δ 13 allele created here was novel. From the FLAG.scpl60 Δ 14 and FLAG.scpl60 Δ 13 templates, we then generated an additional four modified alleles: two in which the deleted KH domains were replaced by the same class of KH domain (i.e., conserved for conserved, or diverged for diverged), and two in which the deleted KH domains were replaced by the opposite class of KH domain. The encoded proteins were designated FLAG.Scpl60 Δ 14.11p, in which conserved

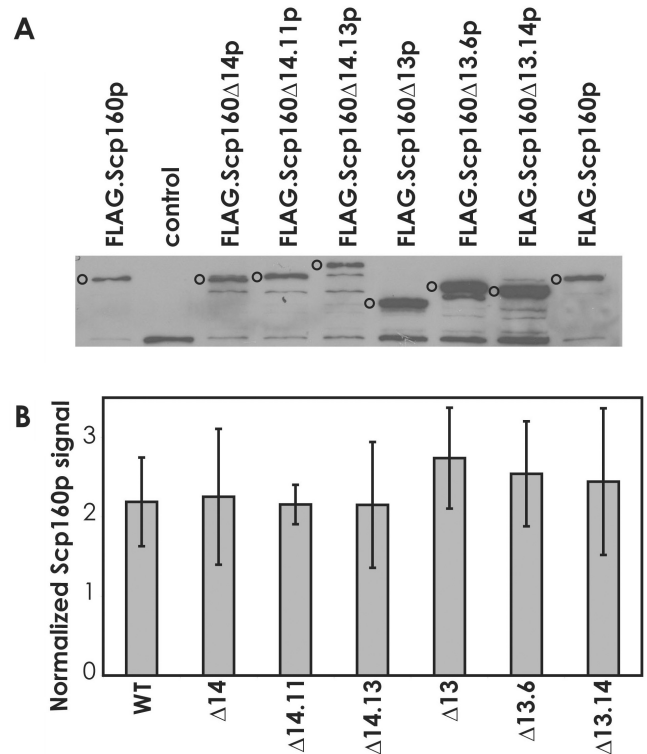


Figure 2. Expression of wild-type and FLAG.Scpl60p variants in yeast. (A) Representative anti-FLAG Western blot analysis of lysates from yeast expressing modified alleles of *scp160*. Open circles mark relevant Scp160p bands, which vary in size, reflecting the indicated KH domain deletions and insertions. Of note, while the size of most KH domains is ~70 amino acids, KH13 is ~140 amino acids in length. The band evident in each lane at the bottom of the panel represents the ~100 kDa endogenous cross-reacting protein that serves as a loading control. (B) Relative abundance of each Scpl60p variant, as defined by scanning densitometry of replicate Western blots, as presented in (A). The FLAG.Scpl60p signal in each lane was normalized to the ~100 kDa cross-reacting band signal in that lane. Values represent means \pm SEM, $n \geq 3$.

KH11 replaced conserved KH14, FLAG.Scpl60 Δ 14.13p, in which diverged KH13 replaced conserved KH14, FLAG.Scpl60 Δ 13.6p, in which diverged KH6 replaced diverged KH13, and FLAG.Scpl60 Δ 13.14p, in which conserved KH14 replaced diverged KH13 (Figure 1). As a result of these manipulations, two of the modified proteins, FLAG.Scpl60 Δ 14p and FLAG.Scpl60 Δ 13p, each carried a total of 13 KH domains, and four of the modified proteins, FLAG.Scpl60 Δ 14.11p, FLAG.Scpl60 Δ 14.13p, FLAG.Scpl60 Δ 13.6p, and FLAG.Scpl60 Δ 13.14p, each carried a total of 14 KH domains. Each modified allele was sequence confirmed and introduced into the appropriate yeast genomic locus in place of the wild-type *SCP160* allele by two-step gene replacement, as described in the Materials and Methods section.

To confirm the expression of each variant Scp160p protein, soluble lysates from mid-log-phase cultures of the appropriate strains were subjected to Western blot analysis with the M2 anti-FLAG monoclonal antibody. Representative results are presented in Figure 2A. Due to the different lengths of the KH domains deleted or

replaced, the protein sizes varied from about 70 amino acids larger than the wild-type protein (FLAG.Scpl60 Δ 14.13p), to about 140 amino acids smaller than the wild-type protein (FLAG.Scpl60 Δ 13p). Yeast expressing untagged Scpl60p served as a negative control for specificity of the antibody, and a \sim 100 kDa endogenous yeast protein that cross-reacts with the M2 antibody (21,22,25) served as a convenient internal control for loading (Figure 2A, bottom-most band in each lane). As illustrated in Figure 2A (open circles), each modified protein was expressed and displayed the expected migration pattern relative to the wild-type FLAG-tagged protein. Furthermore, quantitation of the Scpl60p signal in each lane, relative to the corresponding loading control, for this and replicate experiments, demonstrated that each modified isoform was expressed at a level indistinguishable from that of the wild-type protein (Figure 2B).

In vivo function of Scpl60p KH domain variants

To test the impact of KH domain loss or substitution on Scpl60p function, we first assessed the ability of each variant allele illustrated in Figure 1 to replace *SCP160* *in vivo* (23). As a control, we analyzed the *SCP160* variants that lacked either conserved KH14 or diverged KH13 in parallel with the other variants. Previous studies from our lab (25) had demonstrated the functional significance of KH14, and previous studies from Seedorf and colleagues (24) had demonstrated the functional importance of KH13 + 14 combined. For genetic testing, each modified *scpl60* allele was subcloned into a plasmid and transformed into *scpl60* Δ *eap1* Δ cells maintained by a *URA3* *SCP160* plasmid. Although *SCP160* is not itself essential in yeast (29), deletion of *SCP160* in combination with deletion of *EAP1* is synthetically lethal (23).

To assess how well each modified *scpl60* allele functioned *in vivo*, we applied a previously validated quantitative complementation assay (23). As explained in Materials and Methods, this assay measures the frequency with which a *scpl60* Δ *eap1* Δ mutant strain loses a wild-type *SCP160* maintenance plasmid in the presence of a modified *scpl60* test plasmid. Results are calculated as the relative percentage of maintenance plasmid loss, and the calculated value gives an indication of the functional capacity of the tested allele. For example, a poorly complementing sequence or empty test backbone will have a very low percentage maintenance plasmid loss, because few if any cells will survive loss of the *SCP160* *URA3* maintenance plasmid. In contrast, strains that carry an *SCP160* test sequence that functions well *in vivo* will have a very high degree of maintenance plasmid loss, because most if not all cells will survive loss of the maintenance plasmid. As explained earlier, because the *URA3* maintenance plasmid was the same in all strains, any issues of plasmid replication efficiency or distribution, independent of *SCP160* sequence, should have cancelled out. Further, given that each of the wild-type and variant *SCP160* test sequences was expressed from the same wild-type *SCP160* promoter and plasmid backbone, and that all of these alleles were demonstrated to express comparable levels of Scpl60p protein when genomic (Figure 2),

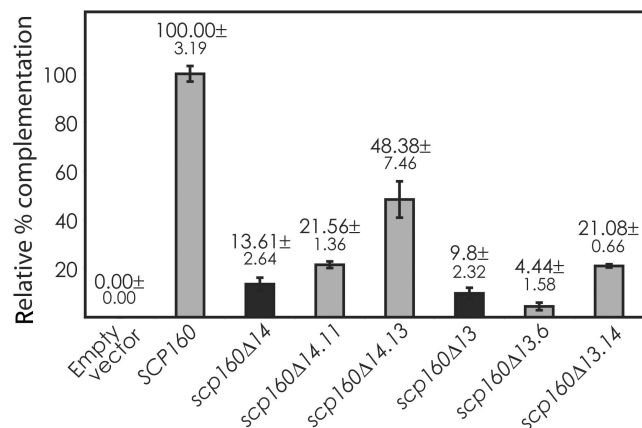


Figure 3. A complementation assay to test the *in vivo* function of *scpl60* variant alleles. As explained in Materials and Methods, the relative functional capacity of any given variant allele of *scpl60* was assessed in terms of its ability to enable survival of a *scpl60* Δ *eap1* Δ double null strain. Results plotted represent the mean of ≥ 3 separate experiments. Error bars represent the standard error of the mean (SEM) for each allele. Values \pm SEM are shown at the top of each bar for the specified allele.

there should have been no disparities in Scpl60p expression levels for this experiment.

Results of these analyses for the control strains confirmed that both KH domains 13 and 14 are essential for Scpl60p function. As illustrated in Figure 3, the *FLAG.scpl60* Δ 14 allele (conserved KH domain deleted) enabled only \sim 14% wild-type levels of maintenance plasmid loss, and the *FLAG.scpl60* Δ 13 allele (diverged KH domain deleted) enabled only \sim 10% wild-type levels of maintenance plasmid loss. While the Δ KH14 allele provided an anticipated result (25), this was the first demonstration that loss of KH13 alone also compromised Scpl60p function. Together, these results provide a foundation for the remainder of the study, which involved testing various add-back alleles for restoration of function.

As illustrated in Figure 3, both of the add-back alleles derived from *FLAG.scpl60* Δ 14, namely *FLAG.scpl60* Δ 14.11, in which a conserved KH11 replaced a conserved KH14, and *FLAG.scpl60* Δ 14.13, in which a diverged KH13 replaced a conserved KH14, demonstrated significant complementation. In fact, *FLAG.scpl60* Δ 14.13 demonstrated the highest degree of complementation of any of the modified *scpl60* alleles tested here. In contrast, while the *FLAG.scpl60* Δ 13.14 add-back allele, in which a conserved KH14 replaced a diverged KH13, demonstrated significant complementation compared to the Δ KH13 variant, the *FLAG.scpl60* Δ 13.6 allele, in which a diverged KH6 replaced a diverged KH13, did not. Indeed, the *FLAG.scpl60* Δ 13.6 allele performed even more poorly than did either deletion variant alone.

Impact of KH domain loss and substitution on macromolecular interactions of Scpl60p

Scpl60p exists as a component of large mRNA/protein complexes (mRNPs) in yeast (21). To assess whether

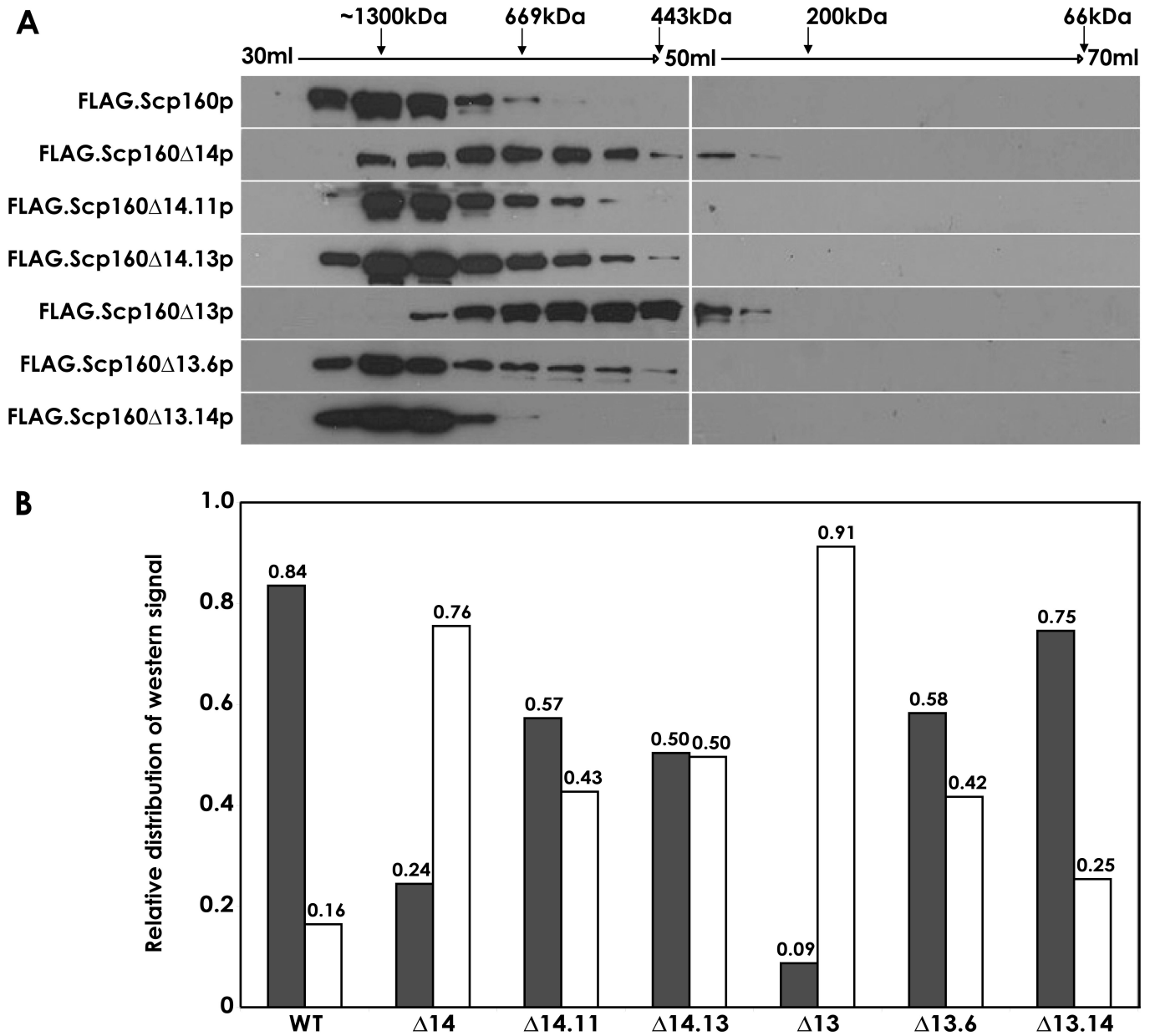


Figure 4. S300 gel filtration chromatography of lysates from yeast expressing Scp160p variants. (A) Lysates from yeast expressing the indicated variants of FLAG.Scip160p were subjected to S300 gel filtration chromatography, as described in Materials and Methods. Fractions were subjected to Western blot analysis using the anti-FLAG M2 antibody. Each panel represents one lysate fractionated into twenty 2-ml samples, as indicated at the top of the figure. The elution positions of the void volume (≥ 1300 kDa) and other size markers are indicated. (B) Signal intensity of each fraction in (A) was quantitated by scanning densitometry and used to calculate the fraction of each variant Scp160p present in samples 1–5 (shaded bars representing the large complexes characteristic of wild-type Scp160p) vs. samples 6–20 (open bars representing smaller complexes). Values plotted are also listed above each corresponding bar.

conserved and diverged KH domains make specific contributions to these interactions, we tested whether the Scp160p variants created here could form mRNPs. For these analyses, soluble lysates prepared from mid-log-phase cultures of yeast expressing each Scp160p variant were size-fractionated over an S300 Sephacryl column, as described in the Materials and Methods section. The resulting fractions were subjected to Western blot analysis using the M2 anti-FLAG antibody to reveal the migration pattern of the Scp160p (Figure 4A) (21,22,25).

Lysates from yeast expressing full-length FLAG.Scip160p were analyzed in parallel as a positive control. To facilitate quantitative comparison among the elution profiles of the different Scp160p variants, the signal intensity in each lane was quantified by scanning densitometry, and the fraction of total signal detected for each protein in the first 5 lanes versus the last 15 lanes is presented in Figure 4B. These studies were replicated at least three times to ensure reproducibility.

As illustrated in Figure 4, wild-type FLAG.Scpl60p migrated through the column predominantly with the void volume, demonstrating that most of the protein existed as part of a large complex (≥ 1.3 mDa). Previous studies have confirmed that these large complexes are mRNPs by virtue of their sensitivity to RNase digestion and the presence of both poly (A) binding protein Pab1p (21) and polyadenylated RNA (18,21,22).

Deletion of either conserved KH14 or diverged KH13 significantly shifted the Scpl60p signal from the void volume to fractions representing smaller complexes, or even monomeric Scpl60p (Figure 4), which comigrates with globular proteins of ~ 450 kDa (21). While both variant migration patterns were clearly different from that observed for wild-type Scpl60p, FLAG.Scpl60 Δ 13p migrated even more aberrantly than did FLAG.Scpl60 Δ 14p (Figure 4B), suggesting that loss of diverged KH13 was even more disruptive to macromolecular interactions than was loss of conserved KH14.

When the missing conserved KH14 sequence was replaced, either by conserved KH11 or by diverged KH13 (Figure 4), close to half of each add-back protein migrated correctly as a large complex, although a 'tail' of smaller complexes extended beyond the normal limit in both profiles. Similarly, when the missing diverged KH13 sequence was replaced, either by diverged KH6 or by conserved KH14, the majority of signal representing each modified protein again migrated as expected (Figure 4). Indeed, the FLAG.Scpl60 Δ KH13.14p protein showed an S300 elution profile that was almost indistinguishable from that of wild-type Scpl60p (Figure 4B). Thus, the conserved and diverged KH domains studied here appear to be functionally interchangeable to a large extent for protein complex formation.

Impact of KH domain loss or substitution on Scpl60p association with polyribosomes

Scpl60p has been implicated in the cytoplasmic metabolism of specific transcripts (18). Consistent with this presumed function, Scpl60p associates with polyribosomes (21,22,24,25). To examine this more specific aspect of Scpl60p macromolecular interaction in the context of KH domain contribution, we assessed the migration through a sucrose gradient of each Scpl60p variant from Figure 1. Sucrose gradients enable the size separation of large complexes, such as ribosomal subunits, monosomes, and polyribosomes. Western blot analyses of the fractions collected from these gradients revealed the distribution profile for each Scpl60p variant protein relative to the migration pattern of the FLAG.Scpl60p wild-type control. Finally, to test whether co-migration with polyribosomes reflected association of each variant Scpl60p with polyribosomes, as opposed to non-specific aggregation, we pretreated relevant samples with RNase or EDTA, as described below. To facilitate comparison between the different profiles, the FLAG-Scpl60p signal intensity in each lane was quantified by scanning densitometry and plotted as the fraction of total signal migrating in samples 1–4 (complexes smaller

than monosomes) versus fractions 5–11 (monosomes and polysomes).

As expected (21,22,25), a majority (>80%) of the wild-type FLAG.Scpl60p protein migrated with monosomes and polyribosomes near the bottom of the gradient, while more than half of the FLAG.Scpl60 Δ 14p variant protein migrated near the top of the gradient (Figure 5A), indicating a loss of association with polyribosomes. Loss of diverged KH13 produced an even more striking shift of Scpl60p signal from the bottom to the top of the gradient (Figure 5A), reconfirming the functional significance of this domain. Of note, although neither wild-type Scpl60p nor any of the Δ KH13 variant proteins showed any signal in the top-most gradient fraction, representing free proteins and small complexes, all three Δ KH14 proteins did show signal in this fraction. The simplest explanation for this disparity is that the wild-type and Δ KH13 Scpl60p proteins assembled entirely into complexes larger than those found in the first gradient fraction, while the three Δ KH14 Scpl60p proteins each maintained at least a subpopulation that did not.

Add-back of conserved KH11 to the Δ KH14 protein had little if any impact on polyribosome association, although add-back of diverged KH13 restored monosome or polyribosome comigration to more than half of the molecules (Figure 5A). A similar 'partial rescue' was seen following add-back of diverged KH6 to the Δ KH13 protein, such that more than half of the FLAG.Scpl60 Δ 13.6p protein migrated with the denser gradient fractions. Finally, add-back of conserved KH14 in place of the missing diverged KH13 completely normalized the sucrose gradient profile of the variant protein (Figure 5A).

To confirm that the pattern of Scpl60p migration evident in these experiments corresponded to association with ribosomes, samples were either treated with EDTA, which dissociates polyribosomes and monosomes into small and large ribosomal subunits, or RNase, which dissociates polyribosomes into monosomes. Both treatments caused the signals for all Scpl60p variants tested to migrate predominantly at the top of the gradient (Figure 5B and C), confirming that the pattern in the absence of pretreatment was due largely to association with polyribosomes, and not simply to aggregation. For all three variant proteins, however, especially Scpl60 Δ 13.6p, a portion of the signal did remain in the denser fractions even after treatment with EDTA or RNase (Figure 5B and C), indicating that these proteins may have aggregated to some extent. By extension, these results suggest that while the data presented in Figure 4 reflect predominantly mRNP formation by the variant Scpl60p proteins, some degree of aggregation may also have occurred.

DISCUSSION

The goal of this study was to explore whether diverged KH domains are as important for Scpl60p function as conserved domains, and also whether diverged KH domains can functionally replace conserved KH domains in Scpl60p. Our results yielded two important conclusions. First, diverged KH domains *can* be as critical for function as conserved KH domains. Specifically,

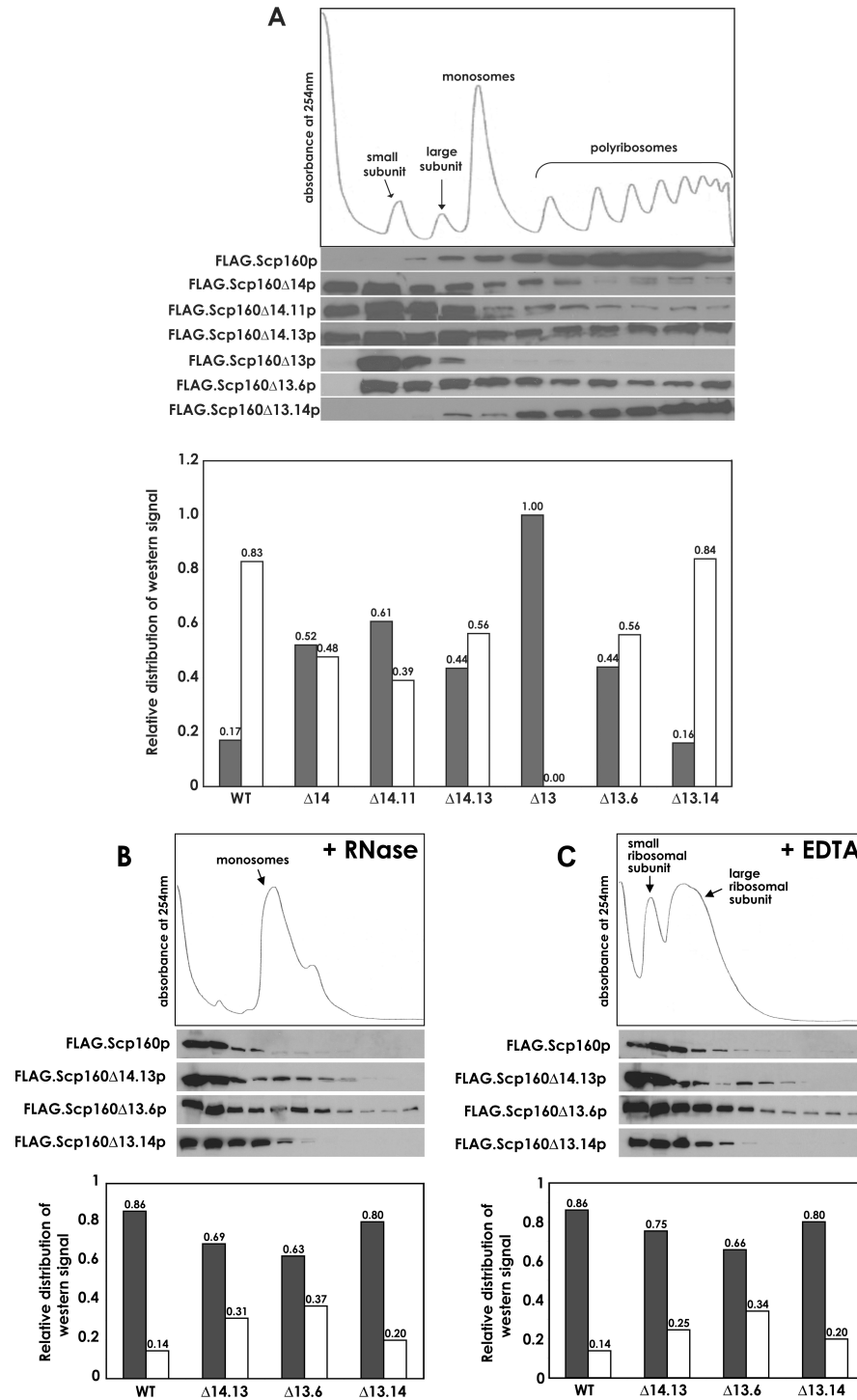


Figure 5. Sucrose gradient fractionation of Scp160p-containing complexes. (A) The upper panel illustrates a representative sucrose gradient profile monitored by absorbance at 254 nm, indicating positions of the small and large ribosomal subunits, monosomes, and polyribosomes. A_{254} profiles of lysates derived from yeast expressing wild-type and Scp160p variants were indistinguishable. The lower seven panels in (A) present anti-FLAG Western blot analyses of gradient fractions from yeast expressing each of the indicated Scp160p variants. Fractionation experiments were repeated at least three times for each strain, with indistinguishable results. The bar graph presents a quantitative assessment of the data illustrated above. In particular, signal intensity of each sample was quantified by scanning densitometry and used to calculate the fraction of each variant Scp160p present in samples 1–4 (shaded bars representing proteins migrating with complexes smaller than monosomes) versus samples 5–11 (open bars representing Scp160p proteins migrating with monosomes or polysomes). Values plotted are also listed above each corresponding bar. (B) Impact of RNase pretreatment on migration of wild-type and variant Scp160p proteins through a sucrose gradient. (C) Impact of EDTA pretreatment on migration of wild-type and variant Scp160p proteins through a sucrose gradient.

we demonstrated, as others have hypothesized (24), that at least one diverged KH domain in Scp160p, KH13, is essential for Scp160p function, despite its lack of the GXXG motif. Second, there can be considerable functional overlap between diverged and conserved KH domains. In particular, we observed that add-back of the diverged KH13 domain sequence to an Scp160p variant missing conserved KH14 partially restored both presumed mRNP formation and polyribosome association, and also restored close to 50% wild-type complementation. It is also important to note, however, that no 'add-back' variant was able to restore fully wild-type levels of *SCP160* function when measured by our biochemical or quantitative complementation assays. Furthermore, the variants studied here showed differing levels of biochemical function with regard to presumed mRNP formation and polyribosome association. These results highlight the distinction between the shared and unique roles of individual KH domains in a multi-KH-domain protein; though they may be repeated, these units are not simply interchangeable.

Size alone is not the answer

Another noteworthy conclusion from our results is that neither the sheer number of KH domains nor the total protein size defines Scp160p function. For example, each of the add-back alleles included a total of 14 KH domains, yet there was enormous disparity in their degrees of function. Furthermore, as illustrated in Figure 2, the variant Scp160p proteins were not all the same size, even after add-back, because KH domain 13 is almost twice the size of any of the other KH domains. This disparity means that the variants missing KH13 with either a KH6 add-back or a KH14 add-back were both approximately the same size as the variant protein that was simply missing KH14, with no add-back, and yet the function of these proteins differed dramatically (Figures 3–5). Clearly, size alone does not define Scp160p function.

Uncoupling aspects of Scp160p function

One of the interesting observations from the data we present here is that *SCP160* function is not binary; the variants we have created and tested clearly uncouple different aspects of Scp160p function. For example, all four of the add-back alleles we tested restored at least partial ability of Scp160p to form presumed mRNPs, but only three restored at least partial ability of the protein to associate with polyribosomes. Finally, of all the variant alleles tested, only one, *SCP160Δ14.13*, restored significant function as measured using our genetic complementation assay. The simplest interpretation of these results would be that *SCP160* function, as measured by the complementation assay, is a complex variable. An ability to form presumed mRNPs appears to be a prerequisite for polyribosome association, and polyribosome association appears to be necessary but not sufficient for full function *in vivo*. It is also interesting to note that the allele demonstrating the strongest ability to complement *scp160Δ eap1Δ* synthetic lethality (*SCP160Δ14.13*)

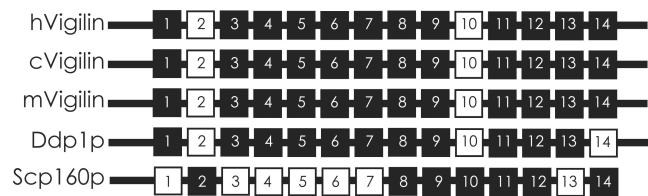


Figure 6. Domain structures of multi-KH-domain-containing proteins. Domain portions of the proteins are not drawn to scale. Numbered, filled boxes represent conserved KH domains; numbered, open boxes represent diverged KH domains. hVigilin: human vigilin; cVigilin: chicken vigilin; mVigilin: mouse vigilin; Ddp1p: *D. melanogaster* multi-KH-domain protein; Scp160p: yeast multi-KH-domain protein. Solid line represents N-terminal and C-terminal protein sequence not predicted to contain KH domain sequences.

was not the same allele that showed the greatest association with polyribosomes (*SCP160Δ13.14*). Furthermore, the most genetically functional of all of the add-back alleles tested was *SCP160Δ14.13*, in which a conserved KH domain was replaced by a diverged KH domain. These data clearly demonstrate that the GXXG motif cannot be essential for the function of every KH domain in Scp160p.

Scp160p and the vigilins

Here we have used Scp160p in yeast as a model to study conserved and diverged KH motifs within a multi-KH-domain protein. Scp160p is most closely related to a family of proteins in higher eukaryotes, known as vigilins (9). First identified in chicken (30), vigilin homologues have now been found in species ranging from *Neurospora crassa* to humans (9,17,19,21,22,30–32)(GenPept 7493335; GenPept 7899383). Like Scp160p, these proteins contain both conserved and diverged KH domains, and as with Scp160p, the diverged KH domains of vigilin have an interrupted GXXG. Interestingly, the *positions* of the diverged KH domains in the human, mouse, and chicken vigilins, and to some extent also in the *Drosophila melanogaster* *DDP1* sequence, have been conserved, but that conservation does not extend to *SCP160* (Figure 6). Nonetheless, fly *DDP1* is apparently adequate to rescue the abnormal DNA-content phenotype in *scp160Δ* yeast (33). While anecdotal, this evidence further supports the conclusion that the presence versus absence of a GXXG motif does not define the functional capacity of every KH domain, at least within the context of a multi-KH-domain protein.

In silico predictions

The KH domain models of two commonly used protein domain databases, SMART and Pfam (34,35), were capable of recognizing all the conserved KH domains in Scp160p, but only a subset of the diverged KH domains, which were not confidently predicted because the scores fell below the allowed significance threshold. Nevertheless, these *in silico* systems were capable of recognizing all the diverged KH domains of most vigilins. It is interesting to note that, while Scp160p KH6 was not recognized by either program, this domain was still able to restore some

biochemical function to the FLAG.Scpl60 Δ 13p variant protein.

By calling into question a disparity between sequence and function among KH domains, our results raise the possibility that there may be more functional KH domains in the proteome than previously appreciated. By extension, these results would suggest that more RNA-binding proteins may exist, as well. Finally, these data clearly imply the need for *in vivo* validation of other sequence motifs defined by *in silico* methods.

ACKNOWLEDGEMENTS

We gratefully acknowledge Brad Bowzard, who helped enormously with the FPLC work, Kim Openo, who helped enormously with the complementation assay, and all other members of the Fridovich-Keil laboratory for their many helpful discussions and contributions. We also thank Ryan Mills for his help with the SMART and Pfam sequence analysis systems, and Ms. Cheryl Strauss for expert editing of the manuscript. This work was supported in part by funds from the National Science Foundation (to JLFK), and in part by funds from the Emory University Research Committee (to JLFK). MB was also supported in part by funds from the National Institutes of Health Training Grant T32 GM-08490 (PI J. Lucchesi). Funding to pay the Open Access publication charge was provided by Emory University.

Conflict of interest statement. None declared.

REFERENCES

- Wells,D.G. (2006) RNA-binding proteins: a lesson in repression. *J. Neurosci.*, **26**, 7135–7138.
- Jong Heon Kim,B.H., Yoon Ki Kim., Mieyoung Choi, and Sung Key Jang., (2000) Protein-protein interaction among hnRNPs shuttling between nucleus and cytoplasm. *J. Mol. Biol.*, **298**, 395–405.
- Moore,M.J. (2005) From birth to death: the complex lives of eukaryotic mRNAs. *Science*, **309**, 1514–1518.
- Chen,Y. and Varani,G. (2005) Protein families and RNA recognition. *FEBS Journal*, **272**, 2088–2097.
- Kielkopf,C.L., Lucke,S. and Green,M.R. (2004) U2AF homology motifs: protein recognition in the RRM world. *Genes Dev.*, **18**, 1513–1526.
- Anantharaman,V. and Aravind,L. (2004) Novel conserved domains in proteins with predicted roles in eukaryotic cell-cycle regulation, decapping and RNA stability. *BMC Genomics*, **5**, 45.
- Siomi,H., Matunis,M.J., Michael,W.M. and Dreyfuss,G. (1993) The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. *Nucleic Acids Res.*, **21**, 1193–1198.
- Adinolfi,S., Bagni,C., Musco,G., Gibson,T., Mazarrella,L. and Pastore,A. (1999) Dissecting FMR1, the protein responsible for fragile X syndrome, in its structural and functional domains. *RNA*, **5**, 1248–1258.
- Currie,J.R. and Brown,W.T. (1999) KH domain-containing proteins of yeast: absence of a fragile X gene homologue. *Am. J. Med. Genet.*, **84**, 272–276.
- Musco,G., Stier,G., Joseph,C., Castiglione Morelli,M.A., Nilges,M., Gibson,T.J. and Pastore,A. (1996) Three-dimensional structure and stability of the KH domain: molecular insights into the fragile X syndrome. *Cell*, **85**, 237–245.
- Baber,J.L., Libutti,D., Levens,D. and Tjandra,N. (1999) High precision solution structure of the C-terminal KH domain of heterogeneous nuclear ribonucleoprotein K, a c-myc transcription factor. *J. Mol. Biol.*, **289**, 949–962.
- Musco,G., Kharrat,A., Stier,G., Fraternali,F., Gibson,T.J., Nilges,M. and Pastore,A. (1997) The solution structure of the first KH domain of FMR1, the protein responsible for the fragile X syndrome. *Nat. Struct. Biol.*, **4**, 712–716.
- Lewis,H.A., Chen,H., Edo,C., Buckanovich,R.J., Yang,Y.Y., Musunuru,K., Zhong,R., Darnell,R.B. and Burley,S.K. (1999) Crystal structures of Nova-1 and Nova-2 K-homology RNA-binding domains. *Structure Fold. Des.*, **7**, 191–203.
- Backe,P.H., Messias,A.C., Ravelli,R.B., Sattler,M. and Cusack,S. (2005) X-ray crystallographic and NMR studies of the third KH domain of hnRNP K in complex with single-stranded nucleic acids. *Structure*, **13**, 1055–1067.
- Paziewska,A., Wyrwicz,L.S., Bujnicki,J.M., Bomszyk,K. and Ostrowski,J. (2004) Cooperative binding of the hnRNP K three KH domains to mRNA targets. *FEBS Lett.*, **577**, 134–140.
- Chmiel,N., Rio,D. and Doudna,J. (2006) Distinct contributions of KH domains to substrate binding affinity of Drosophila P-element somatic inhibitor protein. *RNA*, **12**, 283–291.
- Weber,V., Wernitznig,A., Hager,G., Harata,M., Frank,P. and Wintersberger,U. (1997) Purification and nucleic-acid-binding properties of a Saccharomyces cerevisiae protein involved in the control of ploidy. *Eur. J. Biochem.*, **249**, 309–317.
- Li,A.M., Watson,A. and Fridovich-Keil,J.L. (2003) Scp160p associates with specific mRNAs in yeast. *Nucleic Acids Res.*, **31**, 1830–1837.
- Dodson,R.E. and Shapiro,D.J. (1997) Vigilin, a ubiquitous protein with 14K homology domains, is the estrogen-inducible vitellogenin mRNA 3'-untranslated region-binding protein. *J. Biol. Chem.*, **272**, 12249–12252.
- Frey,S., Pool,M. and Seedorf,M. (2001) Scp160p, an RNA-binding, polysome-associated protein, localizes to the endoplasmic reticulum of Saccharomyces cerevisiae in a microtubule-dependent manner. *J. Biol. Chem.*, **276**, 15905–15912.
- Lang,B.D. and Fridovich-Keil,J.L. (2000) Scp160p, a multiple KH-domain protein, is a component of mRNP complexes in yeast. *Nucleic Acids Res.*, **28**, 1576–1584.
- Lang,B.D., Li,A., Black-Brewster,H.D. and Fridovich-Keil,J.L. (2001) The brefeldin A resistance protein Bfr1p is a component of polyribosome-associated mRNP complexes in yeast. *Nucleic Acids Res.*, **29**, 2567–2574.
- Mendelsohn,B.A., Li,A.M., Vargas,C.A., Riehm,K., Watson,A. and Fridovich-Keil,J.L. (2003) Genetic and biochemical interactions between SCP160 and EAP1 in yeast. *Nucleic Acids Res.*, **31**, 5838–5847.
- Baum,S., Bittins,M., Frey,S. and Seedorf,M. (2004) Asc1p, a WD40-domain containing adaptor protein, is required for the interaction of the RNA-binding protein Scp160p with polysomes. *Biochem. J.*, **380**, 823–830.
- Li,A.M., Vargas,C.A., Brykailo,M.A., Openo,K.K., Corbett,A.H. and Fridovich-Keil,J.L. (2004) Both KH and non-KH domain sequences are required for polyribosome association of Scp160p in yeast. *Nucleic Acids Res.*, **32**, 4768–4775.
- Guthrie,C.A.F., G.R. (1991) Guide to yeast genetics and molecular biology. *Meth. Enzymol.*, **194**, 281–302.
- Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual 2nd edn.* New York, pp. 1.21–1.82, pp. 5.3–6.58.
- Boeke,J.D., Trueheart,J., Natsoulis,G. and Fink,G.R. (1987) 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. *Methods Enzymol.*, **154**, 164–175.
- Wintersberger,U., Kuhne,C. and Karwan,A. (1995) Scp160p, a new yeast protein associated with the nuclear membrane and the endoplasmic reticulum, is necessary for maintenance of exact ploidy. *Yeast*, **11**, 929–944.
- Schmidt,C., Henkel,B., Poschl,E., Zorbas,H., Purschke,W.G., Gloe,T.R. and Muller,P.K. (1992) Complete cDNA sequence of chicken vigilin, a novel protein with amplified and evolutionary conserved domains. *Eur. J. Biochem.*, **206**, 625–634.
- Plenz,G., Kugler,S., Schnittger,S., Rieder,H., Fonatsch,C. and Muller,P.K. (1994) The human vigilin gene: identification, chromosomal localization and expression pattern. *Hum. Genet.*, **93**, 575–582.
- Cortes,A. and Azorin,F. (2000) DDP1, a heterochromatin-associated multi-KH-domain protein of drosophila melanogaster,

- interacts specifically with centromeric satellite DNA sequences [In Process Citation]. *Mol. Cell. Biol.*, **20**, 3860–3869.
33. Cortes,A., Huertas,D., Fanti,L., Pimpinelli,S., Marsellach,F.X., Pina,B. and Azorin,F. (1999) DDP1, a single-stranded nucleic acid-binding protein of *Drosophila*, associates with pericentric heterochromatin and is functionally homologous to the yeast Scp160p, which is involved in the control of cell ploidy. *EMBO. J.*, **18**, 3820–3833.
 34. Finn,R., Mistry,J., Schuster-Bockler,B., Griffiths-Jones,S., Hollich,V., Lassmann,T., Moxon,S., Marshall,M., Khanna,A., Durbin,R. *et al.* (2006) Pfam: clans, web tools and services. *Nucleic Acids Res.*, **34**, Database issue, D247–251.
 35. Schultz,J., Milpetz,F., Bork,P. and Ponting,C. (1998) SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl. Acad. Sci. U. S. A.*, **95**, 5857–5864.